Accounting for the Role of Long Walks on Networks via a New Matrix Function

Ernesto Estrada and Grant Silver

Department of Mathematics & Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XQ, UK

Abstract

We introduce a new matrix function for studying graphs and real-world networks based on a double-factorial penalization of walks between nodes in a graph. This new matrix function is based on the matrix error function. We find a very good approximation of this function using a matrix hyperbolic tangent function. We derive a communicability function, a subgraph centrality and a double-factorial Estrada index based on this new matrix function. We obtain upper and lower bounds for the double-factorial Estrada index of graphs, showing that they are similar to those of the single-factorial Estrada index. We then compare these indices with the single-factorial one for simple graphs and real-world networks. We conclude that for networks containing chordless cycles—holes—the two penalization schemes produce significantly different results. In particular, we study two series of real-world networks representing urban street networks, and protein residue networks. We observe that the subgraph centrality based on both indices produce significantly different ranking of the nodes. The use of the double factorial penalization of walks opens new possibilities for studying important structural properties of real-world networks where long-walks play a fundamental role, such as the cases of networks containing chordless cycles.

Keywords: matrix error functions; matrix tanh function; communicability functions; double-factorial; chordless cycles; complex networks

1. Introduction

The study of large graphs and networks has become an important topic in applied mathematics, computer sciences and beyond [35, 25]. The role played by such large graphs in representing the structural skeleton of complex systems—ranging from social to ecological and infrastructural ones—has triggered the production of many indices that try to quantify the different structural characteristics of these networks [25, 8]. Among those mathematical approaches used nowadays for studying networks, matrix functions [33] of adjacency matrices of graphs have received an increasing visibility due to their involvement in the so-called *communicability functions* [16, 18, 20, 17, 23, 34, 36, 11, 32, 39, 40,

5, 4, 2, 14, 15]. These functions characterize how much information flows between two different nodes of a graph by accounting for a weighted sum of all the routes connecting them. Here, a route is synonymous with a walk connecting two nodes, which is a sequence of (not necessarily distinct) consecutive vertices and edges in the graph. Then, the communicability function between the nodes p and q is defined by the p,q-entry of the following function of the adjacency matrix (see [16, 23] and references therein)

$$G = \sum_{k=0}^{\infty} c_k A^k, \tag{1.1}$$

where the coefficients c_k are responsible of giving more weight to shorter walks. The most popular of these communicability functions is the one derived from the scaling of $c_k = \frac{1}{k!}$, which gives rise to the exponential of the adjacency matrix (see further for definitions). This function, and the graph-theoretic invariants derived from it, have been widely applied in practical problems covering a wide range of areas. Just to mention a few, the communicability function is used for studying real-world brain networks and the effects of diseases on the normal functioning of the human brain [9, 10]. On the other hand, the so-called subgraph centrality [17]—a sort of self-communicability of a node in a graph—has been used to detect essential proteins in protein-protein interaction networks [26, 24]. The network bipartivity—a measure derived from the use of the self-communicability—has found applications ranging from detection of cracks in granular material [38], to the stability of fullerenes [13], and transportation efficiency of airline networks [22].

A typical question when studying the structural indices derived from (1.1) when using $c_k = \frac{1}{k!}$ is whether or not we are penalizing the longer routes in the graph too heavily (see Preliminaries for formal definitions) [21]. To understand this problem let us consider the communicability function between the nodes p and q in the graph:

$$G_{pq} = \sum_{k=0}^{\infty} c_k \left(A^k\right)_{pq}, \tag{1.2}$$

where $\left(A^k\right)_{pq}$ gives the number of routes of length k between these two nodes. Then, when we use $c_k = \frac{1}{k!}$ a route of length 2 is penalized by $1/2$ and a walk of length 3 is penalized by $1/6$. However, a walk of length 5 is already penalized by $1/120 \approx 0.008$, which could be seen as a very heavy penalization for a relatively short walk between these two nodes. This means that the longer walks connecting two nodes make a little contribution to the communicability function. If we consider the function accounting for the self-returning walks starting (and ending) at a given node G_{pp}, a heavy penalization of longer walks means that this index is mainly dependent on the degree of the corresponding node, i.e., the number of edges incident to it. That is,

$$G_{pp} = 1 + c_2 k_p + \sum_{k=3}^{\infty} c_k \left(A^k\right)_{pp}, \tag{1.3}$$

where k_p is the degree of the node p. Then, the main question here is to study whether using coefficients c_k that do not penalize the longer walks as heavily will reveal some structural information of networks which is important in practical applications of these indices.

Here we consider the use of a double-factorial penalization $1/k!!$ [31] of walks as a way to increase the contribution of longer walks in communicability-based functions for graphs and real-world networks. The goal of this paper is two-fold. First, we want to investigate whether this new penalization of walks produces structural indices that are significantly different from the ones derived from the factorial penalization. The other goal is to investigate whether the information contained in longer walks is of significant relevance for describing the structure of graphs and real-world networks. While in the first case we can obtain analytical results that account for the similarities and differences among the two penalization schemes, in the second case we need to use some kind of indirect inference. That is, we aim to explore some practical applications of the indices derived from these two schemes and show whether or not there are significant advantages when using one or the other for solving such practical problems. In the current work we have strong evidences that the contributions of long walks in networks is very important for such graphs containing chordless cycles—also known as holes. In particular we have considered a centrality index based on single- as well as on double-factorial penalization of the walks, and observed that there are significant differences in the ranking of the nodes when the graphs contain such kind of topological features. Chordless cycles are ubiquitous in certain scenarios, such as urban street networks and protein residue networks, which are both studied here. In addition, these chordless cycles are undesired features in certain networks like sensor networks, mobile phone networks and other communication systems, where they represent zones of no coverage of the signals.

2. Preliminaries

We consider in this work simple, undirected and connected graphs $G = (V, E)$ with n nodes (vertices) and m edges. A *walk* of length k in G is a set of nodes $i_1, i_2, \ldots, i_k, i_{k+1}$ such that for all $1 \leq l \leq k$, $(i_l, i_{l+1}) \in E$. A *closed walk* is a walk for which $i_1 = i_{k+1}$. Let A be the adjacency operator on $\ell_2(V)$, namely $(Af)(p) = \sum_{q:\mathrm{dist}(p,q)=1} f(q)$. For simple finite graphs A is the symmetric adjacency matrix of the graph. In the particular case of an undirected network as the ones studied here, the associated adjacency matrix is symmetric, and thus its eigenvalues are real. We label the eigenvalues of A in non-increasing order: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Since A is a real-valued, symmetric matrix, we can decompose it as

$$A = U\Lambda U^T, \tag{2.1}$$

where Λ is a diagonal matrix containing the eigenvalues of A and U is the matrix containing the orthonormalized eigenvectors $\overrightarrow{\psi}_i$ associated with λ_i as its columns. The graphs considered here are connected, therefore A is irreducible and from the Perron-Frobenius theorem we can deduce that $\lambda_1 > \lambda_2$ and that the leading eigenvector $\overrightarrow{\psi}_1$, which will sometimes be referred to as the *Perron vector*, can be chosen such that its components $\overrightarrow{\psi}_{1,u}$ are positive for all $u \in V$.

The degree of a node is the number of edges incident to that node. The graph density is defined as

$$d = \frac{2m}{n\left(n-1\right)}, \tag{2.2}$$

where m is the number of edges in the graph.

The so-called 'exponential' communicability function [18, 16, 23] is defined for a pair of nodes p and q on G as

$$G_{pq} = \sum_{k=0}^{\infty} \frac{\left(A^k\right)_{pq}}{k!} = \left(\exp\left(A\right)\right)_{pq} = \sum_{j=1}^{n} e^{\lambda_j} \overrightarrow{\psi}_{j,p} \overrightarrow{\psi}_{j,q}. \tag{2.3}$$

The G_{pp} terms of the communicability function characterize the degree of participation of a node in all subgraphs of the network, giving more weight to the smaller ones. Thus, it is known as the *subgraph centrality* of the corresponding node [17]. The global structural index defined by

$$EE\left(G\right) = tr\left(\exp\left(A\right)\right) = \sum_{j=1}^{n} e^{\lambda_j}, \tag{2.4}$$

is known as the *Estrada index* of the graph. The indices have been generalized by the use of a parameter β in the matrix function following the work of [19].

$$G_{pq}\left(\beta\right) = \sum_{k=0}^{\infty} \frac{\left(\beta^k A^k\right)_{pq}}{k!} = \left(\exp\left(\beta A\right)\right)_{pq}. \tag{2.5}$$

3. Double-Factorial Penalization of Network Walks

Let us start this section by recalling what the double-factorial is. Let k be a positive integer, then the double-factorial $k!!$ is defined by

$$k!! = \begin{cases} k(k-2)(k-4)...3.1 & k \text{ odd} \\ k(k-2)(k-4)...4.2 & k \text{ even} \\ 1 & k = -1, 0. \end{cases} \tag{3.1}$$

As a variation of the factorial $k!$ the double-factorial appears very suitable for use as the penalization factor of the number of walks of length k in the definition of communicability functions. Other functions have been used in the past for changing the heavy penalization imposed by the single factorial. For instance, the use of $c_k = \alpha^k$ where $0 < \alpha < \lambda_1^{-1}$, has been used since the introduction of the Katz index in 1953 [34]. It is well-known that in this case the Eq. (2.5) converges to the resolvent of the adjacency matrix, i.e., $\left[(I - \alpha A)^{-1}\right]_{pq}$. Another choice of the coefficient c_k in the Eq. (1.2) is to consider $1/(k-t)!$ for some $t > 0$ [21]. In this case the function (1.2) converges to [21]:

$$\left[A^t \left(I + Ae^A - e^A\right)\right]_{pq}. \tag{3.2}$$

The main problem with the two functions previously mentioned is that they are parametric. In the first case we should select the parameter α that is more appropriate for each individual problem. It should be noticed that for very big networks, where $\lambda_1 \gg 1$, the range of this parameter is very narrow leaving very little choice for its variation. In the second case we also need to select the parameter t for each particular problem. Then, our consideration here is the selection of a penalization which is not as heavy as the single factorial but that does not contain any parameter. To see the main differences and similarities with the other penalization discussed before let us consider the terms $A^k/k!!$, where every walk of length k is penalized by $1/k!!$ and let us compare it with the penalization of $1/k!$, α^k and $1/(k-t)!$. To give a simple example we consider a graph having $n = 10$ nodes and $m = 40$ edges and show in Figure 3.1 the values of $c_k \cdot tr\left(A^k\right)$ for $1 \leq k \leq 300$. As can be seen in Figure 3.1 there are significant difference among the three kinds of penalization of walks. The single factorial, $(k-10)!$ and 0.01^k all display very similar behaviour, with very quick decay for relatively small values of k. The use of 0.1^k shows a smoother decay with the increase of k (notice that the plot is semi-log scale). However, for values $1 \leq k \leq 250$ the double-factorial penalizes the walks less heavily than this modified factorial. As can be seen from this Figure, the double-factorial does not penalize the long walks as heavily as the other penalization coefficients, which may retain some important structural information of graphs and networks. Hereafter, we will concentrate our analysis and comparison between the double and the single-factorial penalization of walks in graphs/networks.

As we are interested in defining matrix functions that allow us to calculate several graph invariants, we start by proving the following result.

Lemma 1. *Let A be the adjacency matrix of a simple graph $G = (V, E)$. Then*

$$\sum_{k=0}^{\infty} \frac{A^k}{k!!} = \frac{1}{2}\left[\sqrt{2\pi} erf\left(\frac{A}{\sqrt{2}}\right) + 2I\right] exp\left(\frac{A^2}{2}\right), \tag{3.3}$$

where I is the identity matrix and $erf(A)$ is the matrix error function of A [37].

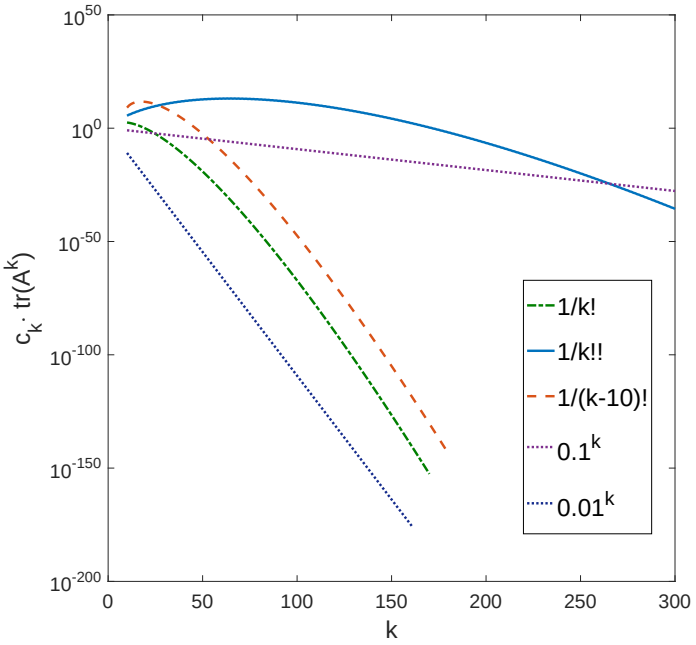

Figure 3.1: Comparison of the decay of $c_k \cdot tr\left(A^k\right)$ for different coefficients c_k (see text for discussion). The y-axis is expressed in logarithmic form.

Proof. Let us consider the spectral decomposition (2.1). Then we can write

$$\begin{aligned}\left(\sum_{k=0}^{\infty}\frac{A^k}{k!!}\right)_{pq} &= \sum_{k=0}^{\infty}\sum_{j=1}^{n}\psi_{j,p}\psi_{j,q}\frac{\lambda_j^k}{k!!}\\ &= \sum_{j=1}^{n}\psi_{j,p}\psi_{j,q}\sum_{k=1}^{\infty}\frac{\lambda_j^k}{k!!}\\ &= \frac{1}{2}\sum_{j=1}^{n}\psi_{j,p}\psi_{j,q}\exp\left(\lambda_j^2/2\right)\left[\sqrt{2\pi}\mathrm{erf}\left(\frac{\lambda_j}{\sqrt{2}}\right)+2\right],\end{aligned}$$

which can be written in the matrix form (3.3), proving the result. □

From the computational point of view the main problem for obtaining (3.3) is provided by the calculation of the matrix error function. In order to circumvent this difficulty we make use here of the remarkable similarity between $\mathrm{erf}(x)$ and $\tanh(x)$ (see Figure 3.2(a)). As can be seen in Figure 3.2(a) there is a gap between the functions in the interval $-2 \leq x \leq 2$. We can definitively improve the similarity between the two function in the following way. The function $[\mathrm{erf}(x) - \tanh(kx)]$ is odd and so its integral from $-\infty$ to ∞ is zero. Then, we will consider the integral

$$\int_0^\infty \left[\mathrm{erf}(x) - \tanh(kx)\right], \tag{3.4}$$

which we will make equal to zero as a way to minimize the difference between the two functions. In other words, we will find the value of k for which (3.4) is zero. Mathematically,

$$\lim_{a\to\infty} \int_0^a \left[\mathrm{erf}(x) - \tanh(kx)\right] = 0,$$

which after integration becomes

$$\lim_{a\to\infty} \left[a\mathrm{erf}(a) - \frac{\ln(\cosh(ka))}{k}\right] = \frac{1}{\sqrt{\pi}}.$$

Using the relation between the hyperbolic cosine and the exponential we have

$$\lim_{a\to\infty} \left[a\mathrm{erf}(a) - \frac{\ln(e^{ka} + e^{-ka})}{k}\right] + \frac{\ln(2)}{k} = \frac{1}{\sqrt{\pi}}$$

As a grows to infinity, e^{-ka} will vanish and $\mathrm{erf}\,(a) = 1$. Then

$$\lim_{a\to\infty} \left[a - \ln\left(e^{ka}\right)\right] + \frac{\ln(2)}{k} = \frac{1}{\sqrt{\pi}},$$

leading to

$$\frac{\ln(2)}{k} = \frac{1}{\sqrt{\pi}}.$$

Which gives us the result of $k = \sqrt{\pi}\ln(2)$. That is, $k = \sqrt{\pi}\ln(2)$ minimizes the gap between the two functions as can be seen in Figure 3.2(b). Consequently, we define the matrix function

$$\begin{aligned} D'\left(A\right) &= \sum_{k=0}^{\infty} \frac{A^k}{k!!} = \frac{1}{2}\left[\sqrt{2\pi}\mathrm{erf}\left(\frac{A}{\sqrt{2}}\right) + 2I\right]\exp\left(\frac{A^2}{2}\right), &(3.5)\\ &\simeq \frac{1}{2}\left[\sqrt{2\pi}\tanh\left(\frac{kA}{\sqrt{2}}\right) + 2I\right]\exp\left(\frac{A^2}{2}\right), &(3.6)\end{aligned}$$

where

$$\tanh(kA) = \frac{e^{kA} - e^{-kA}}{e^{kA} + e^{-kA}}.$$

Hereafter, we define the function

$$D\left(A\right) = \frac{1}{2}\left[\sqrt{2\pi}\tanh\left(\frac{kA}{\sqrt{2}}\right) + 2I\right]\exp\left(\frac{A^2}{2}\right), \tag{3.7}$$

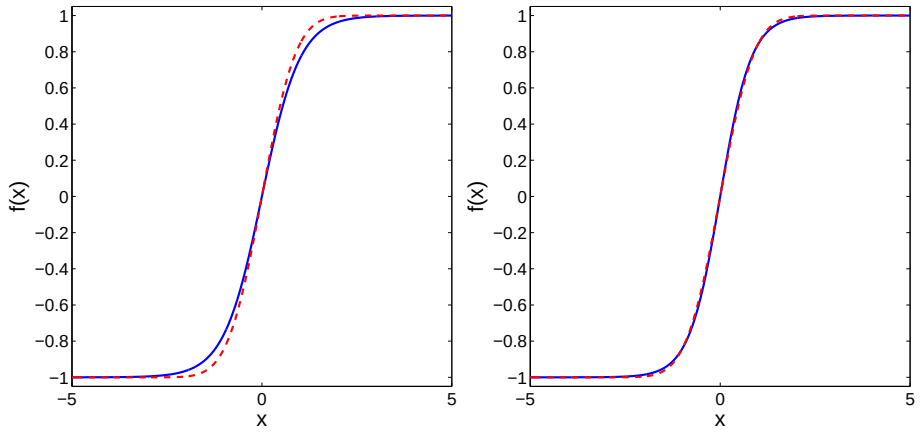

Figure 3.2: (a) Illustration of the similarities between $\mathrm{erf}(x)$ (solid blue line) and $\tanh(x)$ (broken red line). (b) Similar comparison between $\mathrm{erf}(x)$ (solid blue line) and $\tanh(kx)$ (broken red line) for $k = \sqrt{\pi}\log(2)$.

where we use $\tanh(kA)$ instead of $\mathrm{erf}(A)$. It represents an approximate double-factorial or quasi-double-factorial function, but for the sake of simplicity we will simply refer to it generalically as the double-factorial approach. We then define the following indices that will be studied in this work.

Definition 2. Let p and q be any two nodes of the graph G. The double-factorial communicability between these two nodes is defined by $\Gamma_{pq} = (D(A))_{pq}$. Similarly, the term $\Gamma_{pp} = (D(A))_{pp}$ will be called the double-factorial subgraph centrality of the node p and $\Gamma(G) = tr(D(A))$, the double-factorial Estrada index of G.

The generalization of the new matrix function and the indices derived from it lead naturally to considering the following parameter $\beta \in \mathbb{R}$. That is, in general we can consider

$$D(A,\beta) = \frac{1}{2}\left[\sqrt{2\pi}\tanh\left(\frac{k\beta A}{\sqrt{2}}\right) + 2I\right]\exp\left(\frac{\beta^2 A^2}{2}\right), \tag{3.8}$$

and the corresponding indices $\Gamma_{pq}(\beta) = (D(A,\beta))_{pq}$, $\Gamma_{pp}(\beta) = (D(A,\beta))_{pp}$, and $\Gamma(G,\beta) = tr(D(A,\beta))$. Hereafter every time that we write Γ_{pq}, Γ_{pp}, and $\Gamma(G)$ it should be understood that $\beta \equiv 1$.

4. Properties of $\Gamma(G)$

In this section we study some of the mathematical properties of the indices derived from the new matrix function $(D(A))$. In particular, we consider bounds for the double-factorial Estrada index of graphs. In this section we consider that $\beta \equiv 1$, but the results are trivially extended for any β.

Proposition 3. *Let G be a simple connected graph on n nodes. Then, the double-factorial Estrada index of G is bounded as follows*

$$\begin{aligned}\Gamma(G) &\leq \frac{1}{2}\left(\sqrt{2\pi}\tanh\left(\frac{k(n-1)}{\sqrt{2}}\right)+2\right)\exp\left(\frac{(n-1)^2}{2}\right) \\ &+\frac{(n-1)}{2}\left(\sqrt{2\pi}\tanh\left(\frac{-k}{\sqrt{2}}\right)+2\right)\exp\left(\frac{1}{2}\right),\end{aligned} \tag{4.1}$$

with equality if and only if the graph is complete.

Proof. Let l be an edge of G and assume that G is not trivial, i.e., it contains at least one edge. Let $G-l$ be the graph resulting from removing the edge l from G. Let $\mu_k(G)$ be the number of closed walks of length k in G. Then, $\mu_k(G-l)=\mu_k(G)-\mu_k(G:l)$, where $\mu_k(G:l)$ is the number of closed walks of length k in G which contain the edge l. Consequently,

$$\sum_{p=1}^{n}\left(\sum_{k=0}^{\infty}\frac{\mu_k(G-l)}{k!!}\right)_{pp} \leq \sum_{p=1}^{n}\left(\sum_{k=0}^{\infty}\frac{\mu_k(G)}{k!!}\right)_{pp},$$

which means that $\Gamma(G)\leq\Gamma(K_n)$ with equality if the graph is the complete graph won n vertices. We now obtain the formula for $\Gamma(K_n)$. The spectrum of K_n is $\lambda_1=n-1$ with multiplicity one and $\lambda_{j\geq 2}=-1$ with multiplicity $n-1$ from which the result immediately appears. □

Corollary 4. *Let G be a graph and let T be a spanning tree of G. Then*

$$\Gamma(G)\geq\Gamma(T). \tag{4.2}$$

In the next part of this section we will find a lower bound for the double-factorial Estrada index of graphs. First, we find an expression for this index for the path graph P_n, which will be needed for proving the lower bound.

Lemma 5. *Let P_n be a path with n nodes. Then, when $n\to\infty$*

$$\Gamma(P_n)=eI_0(1)\left(n+\frac{1}{2}\right)-\frac{e^2}{2}+o(n), \tag{4.3}$$

where $I_\gamma(z)$ are modified Bessel functions of the first kind [1].

Proof. Let n be number of nodes in P_n

$$\Gamma(P_n) = \sum_{p=1}^{n}\Gamma_{pp}(P_n). \tag{4.4}$$

By substituting the eigenvalues and eigenvectors of the path graph into the expression for Γ_{pp} we obtain

$$\begin{aligned}\Gamma_{pp}(P_n) &= \frac{2}{n+1}\sum_{j=1}^{n}\sin^2\left(\frac{j\pi p}{n+1}\right)\exp\left(2\cos^2\left(\frac{j\pi}{n+1}\right)\right) &(4.5)\\ &= \frac{1}{n+1}\sum_{j=1}^{n}\left[1-\cos\left(\frac{2j\pi p}{n+1}\right)\right]\exp\left(1+\cos\left(\frac{2j\pi}{n+1}\right)\right) &(4.6)\\ &= \frac{e}{n+1}\sum_{j=1}^{n}\left[1-\cos\left(\frac{2j\pi p}{n+1}\right)\right]\exp\left(\cos\left(\frac{2j\pi}{n+1}\right)\right). &(4.7)\end{aligned}$$

Now, when $n\to\infty$ the summation in (4.7) can be evaluated by making use of the following integral

$$\Gamma_{pp}(P_n) = \frac{e}{\pi}\int_0^{\pi}\exp(\cos\theta)d\theta - \frac{e}{\pi}\int_0^{\pi}\cos(p\theta)\exp(\cos\theta)d\theta + o(n), \qquad (4.8)$$

where $\theta = \frac{2j\pi}{n+1}$. Thus, when $n\to\infty$ we have

$$\Gamma_{pp}(P_n) = e\left(I_0(1) - I_p(1)\right) + o(n). \qquad (4.9)$$

We then have

$$\Gamma(P_n) = eI_0(1)n - e\sum_{p=1}^{n} I_p(1) + o(n). \qquad (4.10)$$

Replacing the sum $\sum_{p=1}^{\infty} I_p(x) = \frac{1}{2}\left(e^x - I_0(x)\right)$ we finally obtain the result when $n\to\infty$. □

Now, we can find the lower bound for the double-factorial Estrada index of graphs.

Lemma 6. *Let G be a graph with n nodes. Then,*

$$\Gamma(G) \geq eI_0(1)\left(n+\frac{1}{2}\right) - \frac{e^2}{2}, \qquad (4.11)$$

with equality if and only if G is the path graph P_n, where $I_\gamma(z)$ are modified Bessel functions of the first kind [1].

Proof. Let T be a tree with n nodes

$$\Gamma(T) = \sum_{j=1}^{n} \exp\left(\lambda_j^2/2\right) \tag{4.12}$$

$$= n + \sum_j \lambda_j^2/2 + \sum_j \frac{\left(\lambda_j^2/2\right)^2}{2!} + \sum_j \frac{\left(\lambda_j^2/2\right)^3}{3!} + \cdots \tag{4.13}$$

$$\geq n + m + \frac{m^2}{2} + \frac{m^3}{6} + \frac{m^4}{24} = n + \sum_{k=1}^{4} \frac{(n-1)^k}{k} = F(n) \tag{4.14}$$

where $m = n - 1$ is the number of edges in the path graph. It is easy to show that for $n \geq 3$

$$\Gamma(T) \geq F(n) \geq eI_0(1)\left(n + \frac{1}{2}\right) - \frac{e}{2}^2 = \Gamma(P_n). \tag{4.15}$$

Using Corollary 4 we easily see that $\Gamma(G) \geq \Gamma(T) \geq \Gamma(P_n)$, which proves the result. □

In closing, the double-factorial Estrada index of graphs is bounded $\Gamma(P_n) \leq \Gamma(G) \leq \Gamma(K_n)$, which is similar to $EE(G)$. In the next section we will see that the two indices display significant differences when used to analyze graphs containing significantly large chordless cycles or holes.

5. Graphs with holes

We now consider a graph G and and two nodes p and q in G. Suppose that all the number of subgraphs of sizes smaller than a certain value k_0 to which the node p belongs to is larger than that for the node q. Also consider that the node q is involved in a larger number of subgraphs of size larger than k_0 than the node p. This situation can be found in any graph containing holes. A hole is a chordless cycle, that is a closed sequence of nodes in G such that each two adjacent nodes in the sequence are connected by an edge and each two non-adjacent nodes in the sequence are not connected by any edge in G. Then, the situation previously described can appear when one of the nodes is in a chordless cycle and the other not. An example of this situation is represented in Figure 5.1. Here node A takes place, for instance in 3 triangles, while node B takes place in 6. However, the number of walks of length larger than 17 is bigger for node A, which indeed is part of a chordless cycle of length 18, than for the node B.

Now, let us express mathematically the situation that we have described in the precedent paragraph. That is, $\left(A^k\right)_{pp} > \left(A^k\right)_{qq}$ for all $k < k_0$, and $\left(A^k\right)_{pp} < \left(A^k\right)_{qq}$ for all $k > k_0$, where $k_0 \gg 1$. Let us now consider the difference between the double factorial subgraph centrality and the single-factorial version of it for the node p,

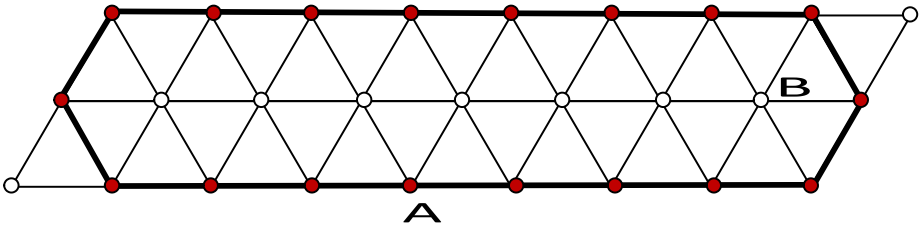

Figure 5.1: Illustration of a triangular lattice with 27 nodes. The nodes marked in red belong to a chordless cycle of length 18, which is highlighted with bolded edges. The nodes A and B are discussed in the main text.

$$\begin{aligned} \Delta_p &= \Gamma_{pp}(G) - EE_{pp}(G) && (5.1)\\ &= \frac{1}{6}\left(A^3\right)_{pp} + \frac{1}{12}\left(A^4\right)_{pp} + \frac{7}{120}\left(A^5\right)_{pp} + \frac{7}{360}\left(A^6\right)_{pp} + \cdots && (5.2)\\ &= \sum_{k=3}^{\infty} \frac{(k-1)!! - 1}{k!!\,(k-1)!!}\left(A^k\right)_{pp}. && (5.3) \end{aligned}$$

Then, in the situation previously described it is plausible that the functions Δ_p and Δ_q follow similar trends to the spectral moments of the adjacency matrix. That is, $\Delta_p > \Delta_q$ for all $k < k_0$, and $\Delta_p < \Delta_q$ for all $k > k_0$.

For the example illustrated in Figure 5.1 we give in Figure 5.2 the plots of $\left(A^k\right)_{pp}/k!!$ and $\left(A^k\right)_{pp}/k!$ for the nodes labelled as A and B in Figure 5.1, as well as the plot of Δ_A and Δ_B. As can be seen the node A, which is in the chordless cycle, has smaller contribution from small subgraphs than node B, which is outside the hole. However, node A takes place in longer subgraphs, such as its own chordless cycle, than node B and consequently $\Delta_A > \Delta_B$ for $k > 18$.

The consequences of the previous kind of situation is that there is a different ranking of the nodes according to the subgraph centrality (or communicability) based on the single and double-factorial penalization. For instance, according to the single factorial penalization $G_{AA} \approx 9.1134$ and $G_{BB} \approx 14.6272$. That is, the node B is more central than node A.. However, according to the double-factorial penalization we obtain $\Gamma_{AA} \approx 3038.6$ and $\Gamma_{BB} \approx 2806.8$, which indicates that indeed the node A is more central than node B. In many cases this difference in ranking is observed for many pairs of nodes in a network, which produces a lack of correlation between the corresponding parameters. In Figure 5 we illustrate the correlations between the subgraph centralities (left panel) and communicability (right panel) for the nodes and pairs of nodes, respectively, in the network illustrated in Figure 5.1.

The graph previously considered is a triangular lattice, which is a planar graph. The situation previously considered where a chordless cycle appears can

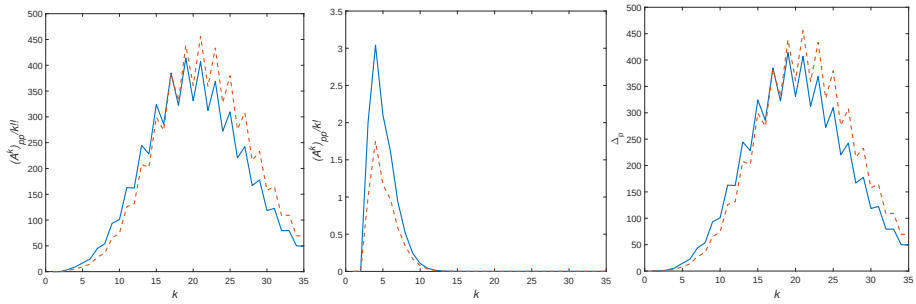

Figure 5.2: Values of $c_k \left(A^k\right)_{pp}$ for $c_k = 1/k!$ and $c_k = 1/k!!$ for different values of k in the triangular lattice illustrated in Figure 5.1 for two nodes. Blue (continuous) line is for the node B and red (broken) line is for the node A. The panel on the right illustrates the values of the differences between both types of penalizations for the two studied nodes (see text for details).

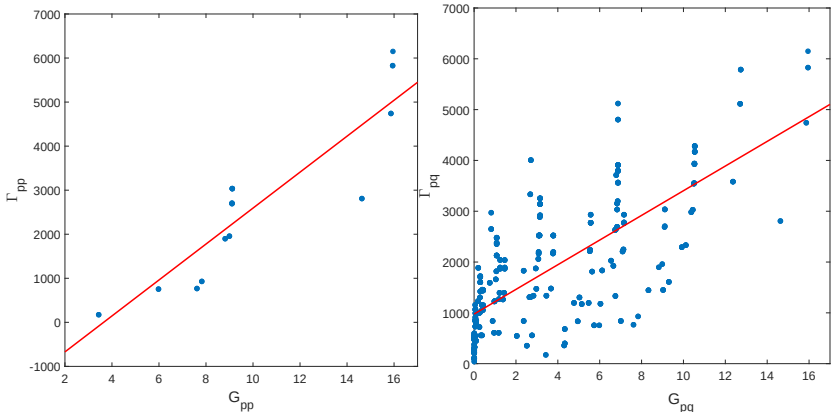

Figure 5.3: Scatterplot of the subgraph centrality (left panel) and communicability (right panel) based on the single and double-factorial penalization for all the nodes and pairs of nodes, respectively, in the graph illustrated in Figure 5.1.

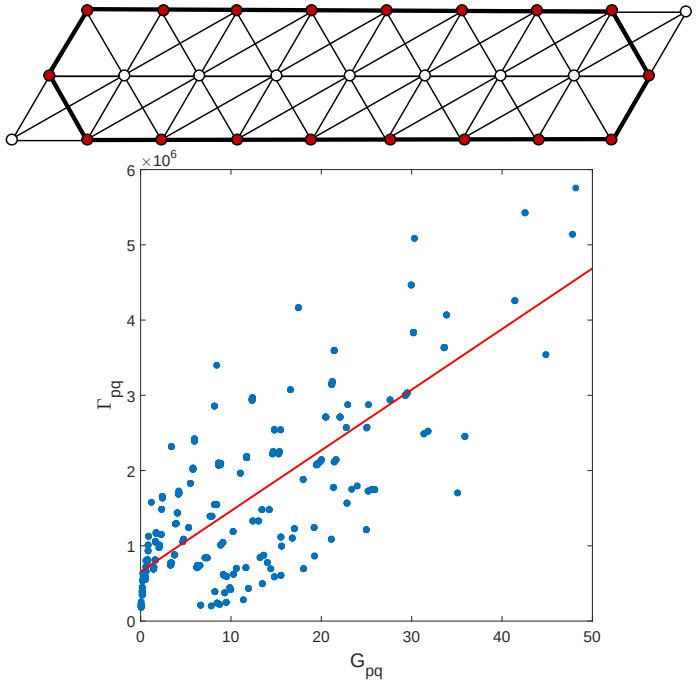

Figure 5.4: (Top) Nonplanar graph obtained from the triangular lattice with 27 nodes illustrated in 5. The nodes marked in red belong to a chordless cycle of length 18, which is highlighted with bolded edges. (bottom) Scatterplot of the communicability based on the single and double factorial for all pairs of nodes for the graph in the top panel.

be seen frequently in these type of graphs and it is consequently of importance for studying certain kinds of real-world networks as we will see in the next section. However, such kinds of examples are not exclusive to planar graphs. As a simple illustration we destroy the planarity of the graph in Figure 5 by adding a few edges but keeping the same chordless cycle as in the original triangular lattice. As can be seen in Figure 5.4 there is a total lack of correlation between the communicability obtained with the single and double factorial for all pairs of nodes in this graph. We will show more realistic examples from real-world networks in the next section of the paper.

6. Analysis of Real-World Networks

6.1. General analysis

An important problem to be considered in practical applications is that the entries of $\frac{A^k}{k!!}$ grow very fast with k in large graphs with relatively high density. Although most real-world networks are sparse, the calculation of indices based on $D(A)$ can be affected by the presence of these very large numbers. For instance, for a network representing the synaptic connections among the neurons

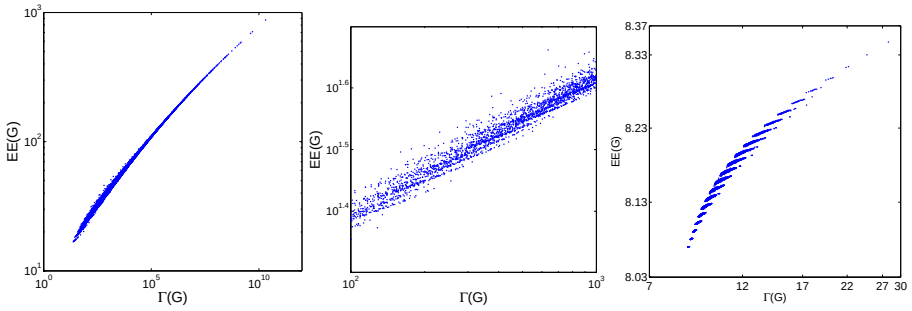

Figure 6.1: (a) Scatterplot of the indices $EE(G,\beta)$ and $\Gamma(G,\beta)$ for all the 11,117 connected graphs on 8 nodes using $\beta = 1$. (b) Magnified plot of the region $100 \leq \Gamma(G) \leq 1000$ for the same plot as in (a). (c) The same as in (a) but using $\beta = 0.1$.

of the worm *C. elegans*, which has $n = 280$ nodes and edge density $d = 0.0505$ the entries of $D(A)$ are bigger than 10^{110}, which far exceeds the largest finite floating-point number in IEEE single precision (10^{38}), but is still below the largest finite floating-point number in IEEE double precision (10^{308}). However, for the network representing the USA system of airports having $n = 332$ nodes and $d = 0.0387$ the entries of $D(A)$ exceed this maximum floating number and a program like Matlab® returns infinity for all its entries. In those cases the adjacency matrix can be multiplied by $\beta < 1$ in order to reduce the magnitude of the entries of $D(A)$ as we will illustrate in some of the examples in this section. We then study the influence of this parameter β on the function $D(A)$.

In this subsection we study networks that do not contain significantly large chordless cycles. We now conduct a computational study of the index $\Gamma(G,\beta)$ of all connected graphs with $n = 4, 5, 6, 7, 8$ nodes and compare it with the index $EE(G,\beta)$ for values $0 < \beta \leq 1$. In Figure 6.1(a) we illustrate the correlation between the two indices for $\beta = 1$ that show the existence of a power-law relation between them. However, by zooming into the smallest valued region of the indices—this region corresponds to graphs with relatively low density of edges—it is revealed that such a correlation between the two indices is far from being simple (see (6.1(b)). This reveals the fact that decreasing the penalization of the walks in graphs from the factorial to the double-factorial make non-trivial changes in the ordering of the graphs, particularly for graphs with relatively low density of edges. This is very important as most real-world networks are sparse and we should expect significant differences between the two different indices for them. More interestingly, we plot the two indices for $\beta = 0.1$ in (6.1(c)) where it can be observed that the correlation between the two indices have now been dramatically decreased.

In order to understand this decay in the correlation between the two indices we express them in terms of the eigenvalues of the adjacency matrix:

$$EE\left(G,\beta\right)=\sum_{j=1}^{n}\exp\left(\beta\lambda_{j}\right), \tag{6.1}$$

$$\Gamma\left(G,\beta\right)=\sum_{j=1}^{n}\exp\left(\frac{\beta^{2}\lambda_{j}^{2}}{2}\right)+\sqrt{\frac{\pi}{2}}\sum_{j=1}^{n}\tanh\left(\frac{k\beta\lambda_{j}}{\sqrt{2}}\right)\exp\left(\frac{\beta^{2}\lambda_{j}^{2}}{2}\right). \tag{6.2}$$

It is easy to see that when $\beta\rightarrow\infty$ both indices are dominated by the principal eigenvalue (spectral radius) of the adjacency matrix, i.e.,

$$EE\left(G,\beta\rightarrow\infty\right)=\exp\left(\beta\lambda_{1}\right), \tag{6.3}$$

$$\begin{aligned}\Gamma\left(G,\beta\rightarrow\infty\right) &= \exp\left(\frac{\beta^{2}\lambda_{1}^{2}}{2}\right)+\sqrt{\frac{\pi}{2}}\tanh\left(\frac{k\beta\lambda_{1}}{\sqrt{2}}\right)\exp\left(\frac{\beta^{2}\lambda_{1}^{2}}{2}\right) &(6.4)\\ &\simeq \left(1+\sqrt{\frac{\pi}{2}}\right)\exp\left(\frac{\beta^{2}\lambda_{1}^{2}}{2}\right), &(6.5)\end{aligned}$$

due to the fact that $\tanh\left(x\right)\approx1$ for $x>5$. Then, it is evident that both indices are highly correlated. On the contrary, when $\beta\rightarrow0$, the second term of (6.2) becomes more relevant. First of all, in this case the term $\tanh\left(x\right)$ is smaller than one for many of the eigenvalues of the network, which means that a larger number of eigenvalues and not only those close to zero make a contribution to this part of the function. Although the first term of (6.2) may still correlate with $EE\left(G,\beta\right)$, the second term does not, which results in a lack of global correlation between $\Gamma\left(G,\beta\right)$ and $EE\left(G,\beta\right)$ when $\beta\rightarrow0$. This lack of correlation for small values of β will be useful for the study of some of the indices derived from the matrix function $D\left(A\right)$.

We now study a group of 61 real-world networks representing social, biological, ecological, infrastructural and technological systems. We first illustrate the correlation between $\Gamma\left(G,\beta\right)$ and $EE\left(G,\beta\right)$ when $\beta=0.2$. To avoid size effects we normalized both indices by dividing their logarithms by the number of nodes. As can be seen in Figure 6.1 in general there is a good correlation between the two indices except for a few networks–about one third of the total number of networks studied—which display large deviations from the linear trend observed. That is, there are 19 networks for which $\left(\log\Gamma\left(G,\beta=0.2\right)\right)/n$ is significantly larger than expected from the linear correlation between this index and $\left(\log EE\left(G,\beta=0.2\right)\right)/n$. Excluding these 19 outliers the Pearson correlation coefficient between the two indices is 0.999. We have calculated the average Watts-Strogatz clustering coefficient and the global transitivity of all the network studied. They have average values for the 61 networks studied of 0.259 and 0.203, respectively. However, if we consider the networks that deviate

significantly from the linear correlation between the two indices, the clustering coefficients have average values of 0.415 and 0.337, respectively, which are much higher than the average observed for the total networks studies. Indeed, if we only consider those networks for which there is a perfect fit between the two indices studied we obtain average clustering coefficients of 0.187 and 0.140, which confirms that the 'anomalous' behavior is observed for networks with the highest clustering coefficients among all the networks studied.

If we consider the difference between the two indices studied we have

$$\Gamma\left(G,\beta\right)-EE\left(G,\beta\right)=\frac{\beta^{3}}{6}tr\left(A^{3}\right)+\frac{\beta^{4}}{12}tr\left(A^{4}\right)+\frac{7\beta^{5}}{120}tr\left(A^{5}\right)+\cdots, \tag{6.6}$$

which clearly indicates that the first term is the one having the largest contribution. We recall that $t=\frac{1}{6}tr\left(A^{3}\right)$, where t is the number of triangles. Then, when $\beta\to 0$ the number of triangles has the largest influence in the difference between the two indices. Consequently, those networks having the largest clustering—which account for the relative abundance of triangles—display the largest difference between the two indices among all the networks studied.

6.2. Centrality

One of the most important uses of matrix functions in the study of networks is the definition of centrality indices. The double-factorial subgraph centrality is defined as the diagonal entries of the matrix function $D\left(A\right)$, which can be expressed in terms of the eigenvalues and eigenvectors of the adjacency matrix as

$$\Gamma_{pp}\left(G,\beta\right)\quad=\sum_{j=1}^{n}\psi_{j,p}^{2}\left[\sqrt{\frac{\pi}{2}}\tanh\left(\frac{k\beta\lambda_{j}}{\sqrt{2}}\right)+1\right]\exp\left(\frac{\beta^{2}\lambda_{j}^{2}}{2}\right). \tag{6.7}$$

The way the double-factorial Estrada index is correlated to $EE\left(G,\beta\right)$ for large values of β is similar to how the double-factorial subgraph centrality $\Gamma_{pp}\left(G,\beta\right)$ is also correlated to the subgraph centrality $G_{pp}\left(G,\beta\right)$ for large values of β. In Figure 6.2 we illustrate the correlations between the subgraph centralities $G_{pp}\left(G,\beta\right)$ and $\Gamma_{pp}\left(G,\beta\right)$ for the protein-protein interaction network of yeast (top plots) and the network of directors in the corporate elite in US (bottom plots). As can be seen in the plots on the left hand side of the Figure, for $\beta=1$ there is a very good linear relation between both centralities as expected from the fact that they can both be approximated by

$$G_{pp}\left(G,\beta\to\infty\right)=\psi_{1,p}^{2}\exp\left(\beta\lambda_{1}\right), \tag{6.8}$$

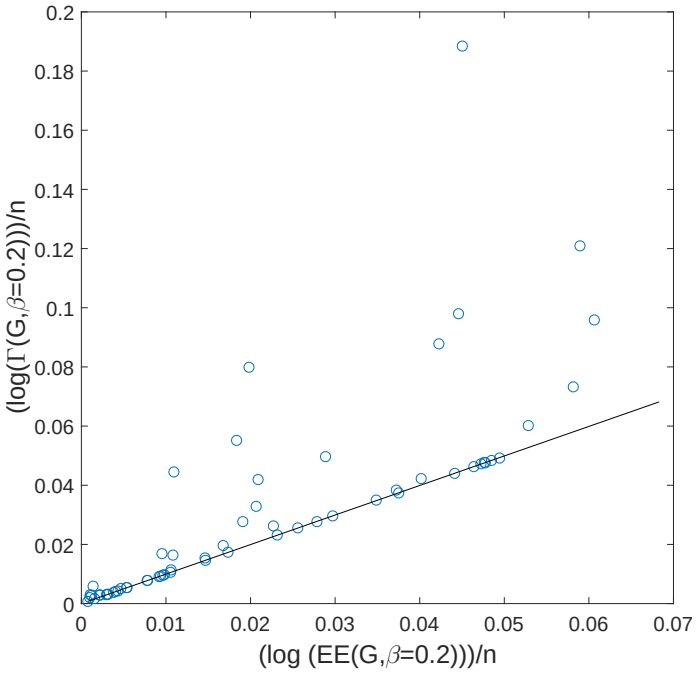

Figure 6.2: Correlation between the logarithms of $\Gamma(G, \beta = 0.2)$ and $EE(G, \beta = 0.2)$ normalized by the number of nodes for 61 real-world networks.

$$\Gamma_{pp}\left(G,\beta\to\infty\right) = \psi_{1,p}^{2}\exp\left(\frac{\beta^{2}\lambda_{1}^{2}}{2}\right) \tag{6.9}$$

$$+ \sqrt{\frac{\pi}{2}}\tanh\left(\frac{k\beta\lambda_{1}}{\sqrt{2}}\right)\exp\left(\frac{\beta^{2}\lambda_{1}^{2}}{2}\right) \tag{6.10}$$

$$\simeq \left(1+\sqrt{\frac{\pi}{2}}\right)\psi_{1,p}^{2}\exp\left(\frac{\beta^{2}\lambda_{1}^{2}}{2}\right). \tag{6.11}$$

However, when $\beta \to 0$, as in the right hand side plots of Figure 6.2, this correlation disappears and both indices differ significantly for several nodes in these networks. The reason for this difference is analogous to the one explained in the previous section for the corresponding Estrada indices.

In order to study how significant the differences between $G_{pp}\left(G,\beta\right)$ and $\Gamma_{pp}\left(G,\beta\right)$ are for relevant network properties we consider the following problem. We consider the identification of essential proteins in the protein-protein interaction network of yeast [26]. Essential proteins are those for which if the corresponding gene is knocked out the entire cell dies. Thus, they are considered to be essential for the survival of the corresponding organisms. In this case we study how many essential proteins exists in the top 10% of proteins ranked according to a given centrality index. The hypothesis behind this experiment is that the most central proteins have higher probability of being essential. Consequently, we rank all the proteins in the yeast PIN according to $G_{pp}\left(G,\beta\right)$ and $\Gamma_{pp}\left(G,\beta\right)$ for $0 \leq \beta \leq 1$ with step 0.01. We then select the top 10% of these proteins and count how many of them are essential. The results for the two centrality indices considered here are illustrated in Figure 6.2(a) where it can be seen that both indices reach the same maximum number of 115 essential proteins identified. However, while $G_{pp}\left(G,\beta\right)$ reaches this maximum for $0.47 \leq \beta \leq 0.57$, the maximum is reached by $\Gamma_{pp}\left(G,\beta\right)$ for $\beta = 0.18$. In the Figure 6.2(b) we illustrate the receiving operating characteristic (ROC) for the classification of the essential proteins using both indices.

Apart from the visual similarities which are evident from a simple inspection of the curves, the quantitative analysis also indicates that there are no significant differences in the quality of the classification using these indices. For instance, the area below the curves for the classification of essential proteins in yeast protein interaction network (PPI) using $G_{pp}\left(G,\beta=0.5\right)$ and $\Gamma_{pp}\left(G,\beta=0.18\right)$ are both 0.69. We can see in (6.2), the indices highly correlate for higher values of β_1,β_2 and also for small β_1,β_2, provided they are close together. In closing, there are no significant differences in the quality of the classification models using $G_{pp}\left(G,\beta\right)$ and $\Gamma_{pp}\left(G,\beta\right)$ when the appropriate values of β are considered. For the sake of comparison we give the values of essential proteins identified by other centrality indices: eigenvector (97); degree (86); closeness (77); betweenness (71) and a random ranking of the proteins identifies 52 essential proteins.

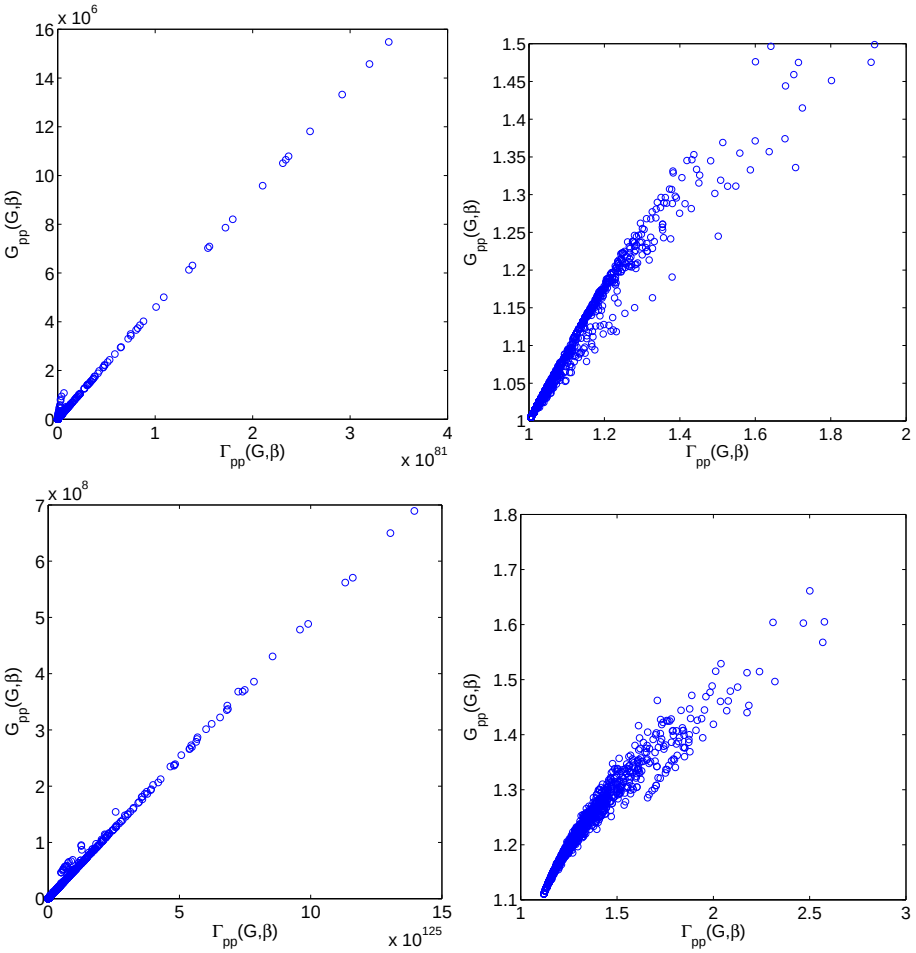

Figure 6.3: Correlation between the subgraph centrality based on the exponential matrix function $G_{pp}(G,\beta)$ and on the matrix function $D(A)$, $\Gamma_{pp}(G,\beta)$ for the protein-protein interaction network of yeast (top plots) and the network of directors in the corporate elite in US (bottom plots). The plots on the left are for $\beta = 1$ and on the right for $\beta = 0.1$.

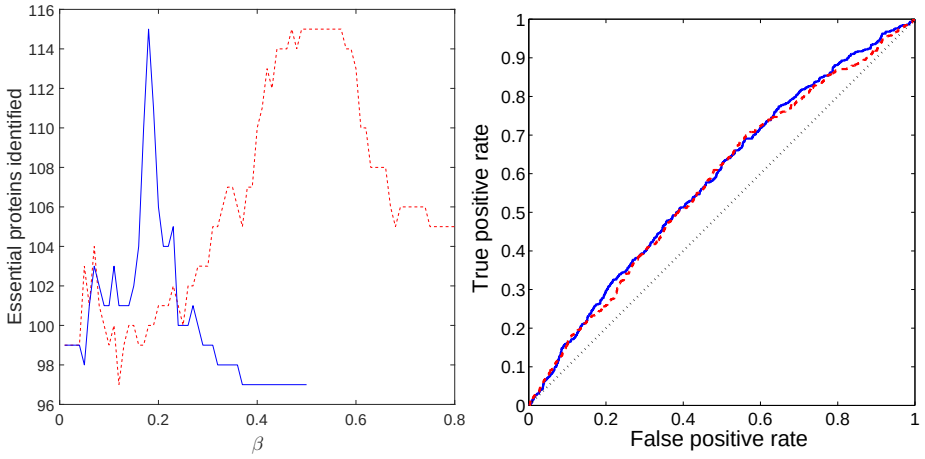

Figure 6.4: (a) Number of essential proteins identified by $G_{pp}(G, \beta = 0.5)$ (red broken curve) and by $\Gamma_{pp}(G, \beta = 0.18)$ (blue continuous line). (b) Illustration of the ROC curves for the classification of essential proteins in yeast protein interaction network (PPI) using $G_{pp}(G, \beta = 0.5)$ (red broken line) and $\Gamma_{pp}(G, \beta = 0.18)$ (blue continuous line).

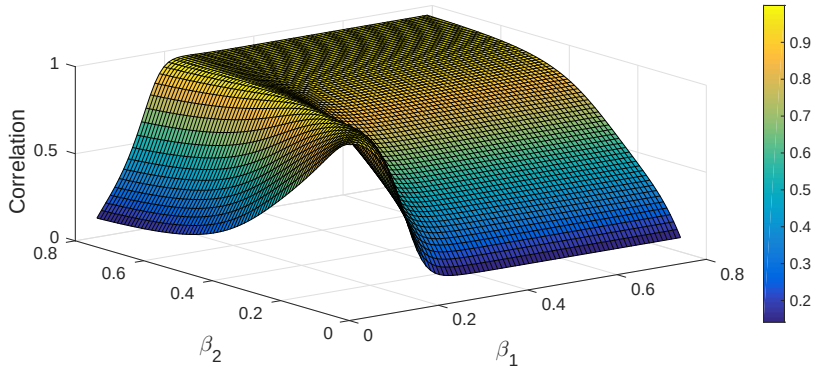

Figure 6.5: Correlation between the diagonal entries of the protein-protein interaction network of yeast based on the exponential matrix function $G_{pp}(G, \beta_1)$ and on the matrix function $D(A)$, $\Gamma_{pp}(G, \beta_2)$ for β_1, β_2 between 0 and 0.75 (step size of 0.01).

The main conclusion of this subsection is that in the case of networks that do not contain significantly large chordless cycles, the double factorial penalization of walks can produce similar results as the indices using single-factorial one when the change of the parameter β is allowed. In the next section we will explore the differences observed between these two schemes for networks containing significantly large holes in their structures.

6.3. Networks with holes

As we have analyzed in Section 5 the graphs containing holes, i.e., chordless cycles, in their structures display a different ranking of the nodes according to the indices developed from the single- and the double-factorial penalization of walks. This situation is frequently observed in real-world networks. A couple of very typical examples are the *urban street networks* [3] and the *protein residue networks* (see for instance [25]). In the first case, the streets of a city are represented as the edges of the network and their intersections are represented by the nodes. In the second case, the nodes represent the α-carbons of the amino acids in the protein and two nodes are connected if the corresponding amino acids are at a distance that allows their physical interaction. In urban street networks—see Figure 6.3—the holes are regions without streets, such as parks, big stores or natural environments like ponds. In the protein residue networks the holes are binding sites—regions in which amino acids are spatially separated to allocate other molecules—for small organic molecules or other proteins. A notable difference between the two systems is that while the first are represented mainly by planar graphs, the second is represented mainly by nonplanar ones. The determination of holes in networks is not a trivial problem and many efforts are directed to this goal due to the importance of these topological features in real-world systems [7, 12, 30]. Information about whether the network contains holes or not can be obtained by means of the so-called "spectral scaling method" [27, 28].

In this Section we compare the ranking of nodes using the subgraph centrality based on single and double-factorial penalization of the walks for a series of urban street networks as well as protein residue networks. For the urban street networks we selected 14 cities from around the World as a representative set. For the protein residue network we selected 14 proteins whose structures have been obtained from x-ray crystallography and deposited in the *Protein Data Bank* (PDB) [6]. Each protein is identified with a unique code of one number and three lowercase letters. A fourth letter sometimes appears to designate the name of the chain to which the protein belongs to. We selected proteins of different domains and sizes ranging from 100 to 1,000 amino acids. In order to compare the ranking based on both approaches, i.e., single- and double-factorial subgraph centralities, we use the *Spearman rank correlation coefficient*. This index indicates how correlated are the ranking of the nodes are according to both indices.

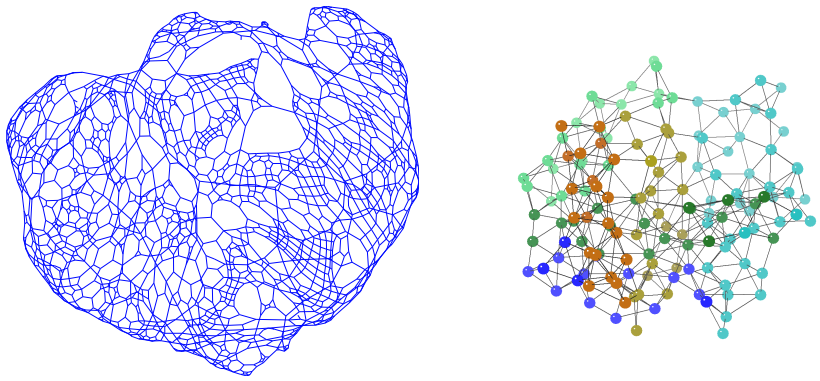

Figure 6.6: Illustration of the urban street network of the city centres of Bologna (left) and a protein residue network (right). Both networks contain chordless cycles (holes) although the first is planar and the second is nonplanar.

In Table 6.3 we give the values of the Spearman rank correlation coefficient for all the networks analyzed in this work. For the urban street networks the average rank correlation coefficient is 0.727 and for the protein residue networks it is 0.747. In both cases the rank correlation indicates that the two indices rank the nodes in significantly different ways. For the sake of comparison the Spearman correlation coefficient between the two indices for the network of Internet as an Autonomous System is 0.9999. This finding—that both indices are not highly rank-correlated—is very important because it indirectly indicates that the double-factorial penalization of walks adds some structural characteristics not described by the single-factorial indices. It is important to remark that in the urban street networks there are rank correlation coefficients as low as 0.56 and as high as 0.92. However, in the protein residue networks the rank correlations are less deviated from the mean. These differences reflect the important fact that cities are in general more heterogeneous than proteins. That is, there are cities with clearly defined holes in their structures while others do not necessarily display such topological features. However, we have previously shown that 95% of representative proteins contain holes indicating that the presence of chordless cycles is a universal property in these systems [29].

In order to illustrate the differences in the ranking of nodes in a network using both types of centrality indices we consider here the urban street network of Cambridge in the UK and the protein residue network of the protein 1qba. In Figure 6.3 we plot the subgraph centralities of each node in the Cambridge urban street network (top images) and of the protein residue network (bottom images) with radius proportional to the logarithm of the subgraph centrality based on single- (left image) and double-factorial (right image). The logarithm is used here to avoid be fooled by the very large numbers of the subgraph

City network	n	Rank Correlation		PDB	n	Rank Correlation
Ahmedabad	4874	0.6002		1ccr	111	0.7296
Atlanta	3234	0.8562		1cpq	129	0.7846
Barcelona	5575	0.9151		1berA	199	0.8215
Berlin	4495	0.6490		1bpyA	326	0.7630
Bologna	825	0.9144		1cem	363	0.8387
Cambridge	1509	0.6513		1chm	401	0.7196
Chengkan	414	0.7127		1bmfA	487	0.7357
Hong Kong	916	0.7417		1ctn	538	0.7272
Mecca	1464	0.6943		1aorA	605	0.7926
Milton Keynes	5581	0.6463		1cyg	680	0.7569
Oxford	1622	0.6951		8acn	753	0.7754
Penang	7055	0.5593		1qba	863	0.6874
Rotterdam	1300	0.6472		1alo	908	0.6494
Yuliang	88	0.8990		1bglA	1021	0.6786

Table 1: Spearman rank correlation coefficients for the subgraph centrality obtained with the single and double-factorial penalization of walks. The results are for the urban street networks (left) and protein residue networks (right) studies here.

centrality in these networks. As can be seen there are significant differences in the centrality of the nodes based on both approaches. The main difference is that the centrality based on the single-factorial identifies fewer hubs than the double-factorial, which is able to delineate complete regions in the networks.

Here we have confirmed what we have first analyzed in Section 5, that in networks containing holes, the subgraph centrality and other indices based on walks are significantly different when the single- or double-factorial penalization are used. In general, we observe significantly different ranking of the nodes, with the subgraph centrality based on double-factorial identifying a larger number of central nodes than the one based on single-factorial penalization. These findings are important for the analysis of specific network problems in particular areas of research, as holes mean different things in different contexts.

7. Conclusion

We have introduced here a new matrix function for studying graphs and real-world networks. This new matrix function of the adjacency matrix of a graph is based on the double-factorial penalization of walks between nodes in a graph. We have observed here that there are two groups of networks for which the behavior of the indices based on the double-factorial penalization changes with respect to that of the single-factorial one. In the first case we have considered networks where there are no structural holes, such as the case

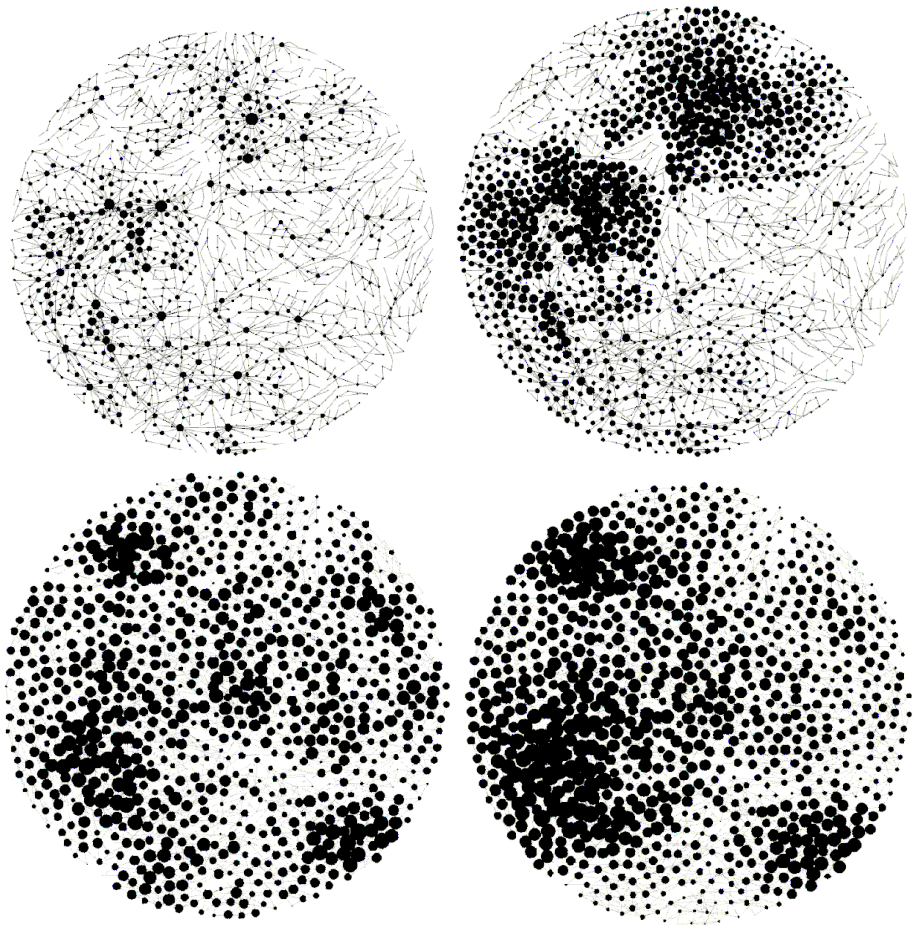

Figure 6.7: Illustration of the subgraph centrality of nodes in the urban street network of Cambridge, UK. The radius of the nodes is proportional to the logarithm of the subgraph centrality based on single- (left image) and double-factorial (right image) penalization of the walks.

of the protein-protein interaction networks of yeast. In this situation we have observed that by introducing a weighting scheme of the form $c_k = \frac{1}{k!}$ or $c_k = \frac{1}{k!!}$ the double factorial indices produce similar results as the single-factorial one for the identification of essential proteins in the yeast PIN.

The second group of networks is formed by those containing significant chordless cycles or holes in their structures, such as urban street networks and protein residue networks. In those cases the contribution of long walks is very important, in particular for navigating around such long holes in the network. In those cases we have shown how a centrality index based on single- as well as on the double-factorial penalization of the walks produce significant differences in the ranking of the nodes. We should stress that significantly large chordless cycles are present in a variety of networks and that their study is of major importance in communication systems, where they should be avoided as regions of zero-coverage of the communication signals. Consequently, the new scheme of penalizing walks by using the double-factorial opens new possibilities for the study of many problems in real-world networks.

8. Acknowledgment

EE thanks the Royal Society of London for a Wolfson Research Merit Award. GS thanks EPRSC for a PhD scholarship. The authors thank both referees for excellent revision and constructive comments on the manuscript.

[1] M. Abramovich, I. Stegun, Handbook of Mathematical Functions with Formulas, graphs and mathematical Tables, National Bureau of Standarts, Appl. Math. series 55 (1964).

[2] F. Arrigo, M. Benzi, Updating and downdating techniques for optimizing network communicability, SIAM J. Sci. Comp. 38 (2016) B25–B49.

[3] M. Barthélemy, Spatial networks, Phys. Rep. 499 (2011) 1–101.

[4] M. Benzi, C. Klymko, On the limiting behavior of parameter-dependent network centrality measures, SIAM J. Matrix Anal. Appl. 36 (2015) 686–706.

[5] M. Benzi, E. Estrada, C. Klymko, Ranking hubs and authorities using matrix functions, Linear Algebra Appl. 438 (2013) 2447–2474.

[6] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[7] N. Chandrasekharan, V. S. Lakshmanan, M. Medidi, Efficient parallel algorithms for finding chordless cycles in graphs, Parallel Proc. Lett. 3 (1993) 165–170.

[8] L. F. Costa, O. N. Oliveira Jr, G. Travieso, F. A. Rodrigues, and P. R. Villas Boas, L. Antiqueira, M. P. Viana, and L. E. Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications, Adv. Phys. 60 (2011) 329–412.

[9] J.J. Crofts, D.J. Higham, A weighted communicability measure applied to complex brain networks, J. Roy. Soc. Interface 6 (2009) 411–414.

[10] J.J. Crofts, D.J. Higham, R. Bosnell, S. Jbabdi, P.M. Matthews, T.E.J. Behrens, H. Johansen-Berg, Network analysis detects changes in the contralesional hemisphere following stroke, Neuroimage 54 (2011) 161–169.

[11] H. Deng, S. Radenković, I. Gutman, The Estrada index, in: D. Cvetković, I. Gutman (Eds.), Applications of Graph Spectra, Math. Inst, Belgrade, 2009, pp. 123–140.

[12] E. S. Dias, D. Castonguay, H. Longo, W. A. R. Jradi, Efficient enumeration of all chordless cycles in graphs, Comput. Res. Repos. abs/1309.1051 (2013).

[13] T. Došlić, Bipartivity of fullerene graphs and fullerene stability, Chem. Phys. Lett. 412 (2008) 336–340.

[14] V. Ejov, J.A. Filar, S.K. Lucas, P. Zograf, Clustering of spectra and fractals of regular graphs, J. Math. Anal. Appl. 333 (2007) 236–246.

[15] V. Ejov, S. Friedlan, G.T. Nguyen, A note on the graph's resolvent and the multifilar structure, Linear Algebra Appl. 431 (2009) 1367–1379.

[16] E. Estrada, D. J. Higham, Network properties revealed through matrix functions, SIAM Rev. 52 (2010) 96–714.

[17] E. Estrada, J.A. Rodríguez-Velázquez, Subgraph centrality in complex networks, Phys. Rev. E 71 (2005) 056103.

[18] E. Estrada, N. Hatano, Communicability in complex networks, Phys. Rev. E 77 (2008) 036111.

[19] E. Estrada, N. Hatano, Statistical-mechanical approach to subgraph centrality in complex networks, Chem. Phys. Lett. 439 (2007) 247–251.

[20] E. Estrada, Characterization of the folding degree of proteins, Bioinformatics 18 (2002) 697–704.

[21] E. Estrada, Generalized walks-based centrality measures for complex biological networks, J. Theor. Biol. 263 (2010) 556–565.

[22] E. Estrada, J. Gómez-Gardeñes, Network bipartivity and the transportation efficiency of European passenger airlines, Physica D 323-324 (2016) 57–63.

[23] E. Estrada, N. Hatano, M. Benzi, The physics of communicability in complex networks, Phys. Rep. 514 (2012) 89–119.

[24] E. Estrada, Protein bipartivity and essentiality in the yeast protein–protein interaction network, J. Proteome Res. 5 (2006) 2177–2184.

[25] E. Estrada, *The Structure of Complex Networks. Theory and Applications*, Oxford University Press, 2011.

[26] E. Estrada, Virtual identification of essential proteins within the protein interaction network of yeast, Proteomics 6 (2006) 35–40.

[27] E. Estrada, Spectral scaling and good expansion properties in complex networks, Europhys. Lett. 73 (2006) 649.

[28] E. Estrada, Topological structural classes of complex networks, Phys. Rev. E 75 (2007) 016103.

[29] E. Estrada, Universality in protein residue networks, Biophys. J. 98 (2010) 890–900.

[30] Q. Fang, J. Gao, L. J. Guibas, Locating and bypassing holes in sensor networks, Mobile Net. Appl. 11 (2006) 187–200.

[31] H.W. Gould, J. Quaintance, Double fun with double factorials, Mathematics Magazine 85 (2012) 177–192, .

[32] I. Gutman, H. Deng, and S. Radenković, The Estrada index: an updated survey, in: D. Cvetković, I. Gutman (Eds.), Selected Topics on Applications of Graph Spectra, Math. Inst., Beograd, 2011, pp. 155–174.

[33] N. J. Higham, *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.

[34] L. Katz, A new index derived from sociometric data analysis, Psychometrika 18 (1953) 39–43.

[35] M. E. J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.

[36] J.A. de la Peña, I. Gutman, J. Rada, Estimating the Estrada index, Linear Algebra Appl. 427 (2007) 70–76.

[37] R. Taylor, Solution of the linearized equations of multicomponent mass transfer, Ind. Eng. Chem. Fundam. 21 (1982) 407–413.

[38] D.M. Walker, A. Tordesillas, Topological evolution in dense granular materials: a complex networks perspective, Int. J. Solids Struct. 47 (2010) 624–639.

[39] J. Wu, M. Barahona, Y-J. Tan, H-Z. Deng, Robustness of regular ring lattices based on natural connectivity, Int. J. Syst. Sci. 42 (2011) 1085–1092.

[40] J. Wu, H.-Z. Deng, Y.-J. Tan, D.-Z. Zhu, Vulnerability of complex networks under intentional attack with incomplete information, J. Phys. A: Math. Theor. 40 (2007) 2665.